CrossMark

ORIGINAL ARTICLE

Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran

Hamid Reza Pourghasemi[1] · Norman Kerle[2]

Abstract In many parts of the world, landslide susceptibility remains inadequately mapped, due to the lack of both data and suitable methods for widespread implementation. Iran is one of those countries with extensive landslide problems, with nearly 4900 large landslides occurring between 1993 and 2007. At the same time landslide susceptibility has not been assessed for the country. Random forest (RF) has recently been shown to be a suitable tool for such mapping. In this study we further coupled the RF method with an evidential belief function (EBF) approach, and tested the suitability for landslide susceptibility mapping for variable terrain and data conditions in the west of Mazandaran Province, northern Iran. Locations of earlier landslides were identified by interpreting aerial photographs and through extensive field surveys. Eleven conditioning factors were used in the RF model. The spatial relationship between landslide occurrence and conditioning factors was then assessed using the data-driven EBF model, and EBF values paired to each map. Finally, the EBF maps were used for running the RF model. Finally, the efficiency of the RF-EBF model was tested using the area under the curve to measure the success and prediction rates of the incorporated data. This resulted in a success rate of 85.2 %, and a prediction rate of 81.8 %. The most important conditioning factors identified were lithology, altitude, distance from roads, and land use, respectively. Based on the overall assessment, the combined RF and EBF approach was found to be objective and an applicable estimator that improves the predictive accuracy and controls for overfitting, and thus useful for landslide susceptibility mapping at regional scales.

Keywords Landslide · Random forest · Evidential belief function · Iran · Machine learning · Slope instability

✉ Hamid Reza Pourghasemi
hr.pourghasemi@shirazu.ac.ir;
hamidreza.pourghasemi@yahoo.com

[1] Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran

[2] Department of Earth Systems Analysis (ESA), Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands

Introduction

Landslides are natural geologic processes that cause different types of damages and effects on human life and human-made structures (Goetz et al. 2011). Mihir et al. (2014) stated that approximately 375 fatal landslides occur globally each year, with around 4600 people killed annually. Previous studies have highlighted the increasing costs of landslide damages, which can be primarily attributed to increased human occupation and activities in vulnerable environments. Landslide susceptibility maps are a prerequisite for planners, decision makers, and engineers in slope management and land use planning for this problem to be effectively addressed. In recent decades, progress has been made in landslide susceptibility mapping because of the rapid development of remote sensing techniques, modeling tools and Geographical Information Systems (GIS) (Weirich and Blesius 2007). For landslide susceptibility mapping, landslide occurrences depend generally on the complex interactions among a large number of conditioning factors. Therefore, investigations into landslide susceptibility require an assessment of the relationships between various terrain conditions and landslide occurrence. An experienced Earth scientist has the capabilities to assess the overall slope

Springer

conditions and to extract the critical parameters. However, an objective procedure that can be automated and applied over large regions is often desired to quantitatively support the slope instability assessment. This procedure requires the evaluation of the spatially varying terrain conditions, as well as the spatial representation of landslides.

Different techniques have previously been used for landslide susceptibility mapping. Many studies have evaluated landslide susceptibility using GIS and frequency ratio models (Pourghasemi et al. 2012a; Nourani et al. 2014; Park and Lee 2014), bivariate and multivariate statistical methods (Pradhan 2010; Lee et al. 2012; Poiraud 2014; Umar et al. 2014; Youssef et al. 2014a; Fiorucci et al. 2015; Youssef 2015), index of entropy (Bednarik et al. 2010; Constantin et al. 2011; Jaafari et al. 2014; Park 2014; Youssef et al. 2014b), analytical hierarchy process (Hasekiogullari and Ercanoglu 2012; Park et al. 2013; Pourghasemi et al. 2014a; Youssef 2015), fractal theory (Li et al. 2012; Pourghasemi et al. 2014b), support vector machine (Pradhan 2013; Li and Kong 2014; Ren et al. 2014; Wu et al. 2014; Su et al. 2015), artificial neural networks (Pradhan and Buchroithner 2010; Song et al. 2012; Wu et al. 2013; Nourani et al. 2014; Chen et al. 2014; Moosavi and Niazi 2015; Wang et al. 2015; Dou et al. 2015), fuzzy approaches (Ercanoglu and Gokceoglu 2004; Akgun et al. 2012; Osna et al. 2014; Bortoloti et al. 2015; Meten et al. 2015), object-based image analysis (Martha et al. 2012), and some hybrid methods, including the neuro-fuzzy model (Sezer et al. 2011; Pradhan 2013; Lee et al. 2015a, b). Recently, some new methods and techniques have been used for landslide susceptibility mapping such as random forest (RF) (Miner et al. 2010; Stumpf and Kerle 2011; Vorpahl et al. 2012; Catani et al. 2013; Micheletti et al. 2014; Youssef et al. 2015) and Bayesian logistic regression (Das et al. 2012). Bivariate statistical models such as frequency ratio are easy to run in a GIS (Oh et al. 2011b), but have a disadvantage that equal weights are assumed for different effective factors. In the weights of evidence (WofE) method, the contrast value is the rating of each class of each factor that influences landslide occurrence. This value is positive for a positive spatial association and negative for a negative spatial association (Lee and Choi 2004). Logistic regression determines the type of relation between the landslide occurrence and effective factors, which can be positive or negative. The index of entropy determines the weight of each effective factor on the occurrence of landslides. ANN needs an optimal network architecture design (number of hidden layers and units) and the setting of the optimal cost and weight decay values, which makes it complex to use (Rodriguez-Galiano and Chica-Rivas 2012). The evidential belief function (EBF) model supports a series of mass functions, including belief, disbelief, uncertainty and plausibility. Thus, the results of the model present the quantitative relationships between landslide occurrence and effective factors by modeling the degree of uncertainty (Pourghasemi and Beheshtirad 2014). Saaty's AHP is a widely used multi-criteria decision making technique in the field of natural resources and environmental management. In general, the RF algorithm has several benefits in comparison with other commonly used multivariate regression or classification methods. First, it does not require prior assumptions about the distribution of the explanatory variables; secondly, it allows for the mixed use of categorical and numerical variables without resorting to the use of indicator (or dummy) variables, and third, it is capable of considering interactions and nonlinear relationships between variables. Although many different models and techniques for landslide susceptibility mapping have been proposed and implemented, no agreement has so far been reached on which model and techniques are the best for landslide susceptibility mapping (Wang and Sassa 2005).

Thus, the main purpose of this research is to determine the relationship between landslide occurrences and conditioning factors, using an EBF model, landslide susceptibility mapping using a GIS-based RF model, and evaluation of the combined method in the Safarood Watershed, Mazandaran Province, northern Iran. The main advantages of RF are that it is one of the most accurate learning algorithms available, and that it runs efficiently on large databases (Lee et al. 2015a, b). The accuracy and variable importance values are generated automatically, an important benefit in landslide susceptibility assessment. Moreover, overfitting is not a problem in the mentioned algorithm, and neither is it very sensitive to outliers in the training dataset (Breiman 2001). Therefore, the RF algorithm, coupled with EBF, can be a good approach for other parts of the country, or even other countries for land use planning and decision making in landslide risk areas.

Study area

The Safarood Watershed is located in the western part of Mazandaran Province, Iran. The considered site lies between the latitudes 36°47′29″N to 36°57′35″N, and the longitudes 50°24′16″E to 50°41′39″E (Fig. 1). It covers an area of approximately 162.6 km^2. The topographical elevation of study area varies between 20 and 3540 m a.s.l. The slope angles of the area range from 0° to as much as 70°. A rainfall map was prepared using the data collected at six rainfall stations (Ramsar, Gavermak, Mianlat, Abe Madani-e-Nidshet, Gardeh Poshtehsar, and Zarodal) for the years 1975–2012 (annual mean rainfall). The mean annual rainfall is around 1,220 mm between September and December, according to the Islamic Republic of Iran
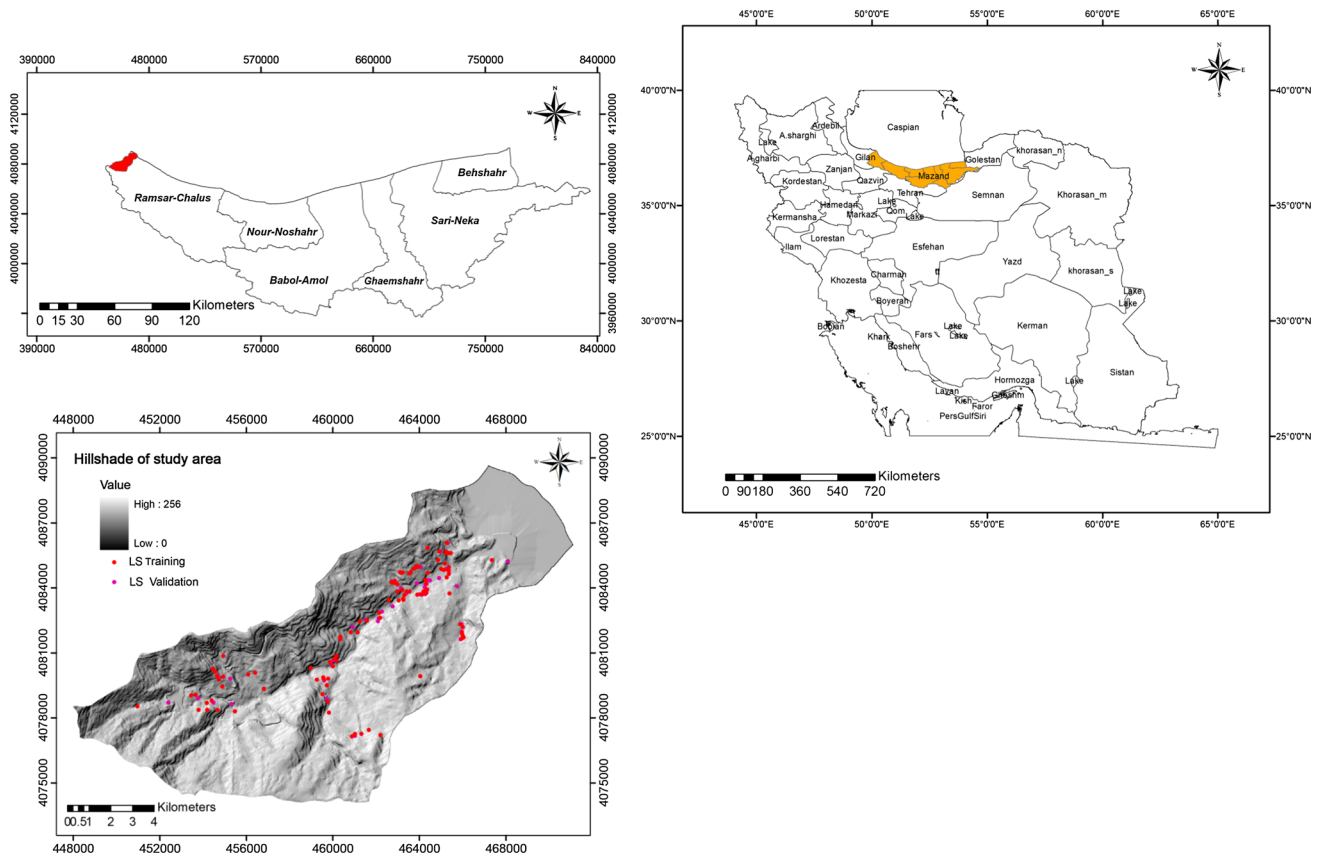
Fig. 1 Map of the study area, with indication of past landslides

Meteorological Organization (IRIMO 2012), whose data also show the temperature in the Safarood area ranging between 21 and 30 °C. The study area is covered by various types of lithological formations, including Quaternary, Jurassic, Triassic, and Late Permian types of formations (Table 1). The Quaternary deposit covers 15.56 % of the study area and includes Q^{al} (recent loose alluviums), Q_2^d (undivided deltaic alluviums), Q_2^m (marine deposits), Q2t1 g (old gravelly terrace), Q_2^s (scree and rock falls), and $Q_{2\ 2}^{tg}$ (young gravelly terrace). The j1^c-2 (dark gray polymictic conglomerate with abundant quartzite pebbles, with inter layers of siltstone, sandstone and coal), TR_e^{d1} (light gray-to-cream dolomite) and La (laterite) covers about 80 % of the study area (GSI 1997). Its folded mountain system regionally strikes northwest to the southeast. The Binaksar fault is one of the main faults in the study area. Most of the area (64 %) is covered by forest. Other land use classes are dry-farming and forest, forest, forest and orchard, irrigation agriculture, irrigation agriculture and orchard, orchard, range, tea land, and residential area. Generally, mountainous features, high tectonic activity, and geological and climatologically variety caused to the Iranian plateau being susceptible to various kinds of landslides, especially in the Alborz active

mountainous belts. Alborz is a mountain range in northern Iran that stretches from the border of Azerbaijan along the western and entire southern coast of the Caspian Sea, and finally runs northeast and merges into the Aladagh Mountains in the northern parts of Khorasan (https://en.wikipedia.org/wiki/Alborz). According to earlier studies by the Iranian Landslide Working Party (ILWP 2007), the highest frequency of landslide occurrence in Iran is in the Mazandaran Province. The results showed that more than 520 landslides occurred between 2005 and 2013 and caused extensive damages. Evidence for that includes the destruction of some 438 ha forest lands, 1264 ha agriculture lands and gardens, 12 km roads, as well as more than 50 villages (http://irna.ir/fa/NewsPrint.aspx?ID=80365819). The Safarood Watershed is one of the most important ecotourism areas in Iran and the Mazandaran Province, because of its climate conditions, geomorphological attributes, and ecological landscape. Despite the mentioned potential (important ecotourism area), deforestation, road construction, land use changes, and expansion of villas along high slopes have led to increasing landslide occurrence in the study area. The area is highly prone to landslide and this susceptibility needs to be assessed for a safer and more optimized use of the land.

Table 1 Types of geological formations in the study area

Class	Code	Formation	Geological age	Lithology
A	Q^{al}	–	Quaternary	Recent loose alluviums
B	Q_2^d	–	Quaternary	Undivided deltaic alluviums
C	Q_2^m	–	Quaternary	Marine deposits
D	Q2t1 g	–	Quaternary	Old gravelly terrace
E	Q_2^s	–	Quaternary	Scree and rock falls
F	Q_2^{tg}	–	Quaternary	Young gravelly terrace
G	$J_{2\text{-}3}^{1,dl}$	–	Jurassic	Light gray-to-cream, medium-to-thick-bedded limestone and dolomite
H	$J_{1\text{-}2}^c$	–	Jurassic	Dark gray polymictic conglomerate with abundant quartzite pebbles, with interlayers of siltstone, sandstone and coal
I	Jk^1	–	Jurassic	Gray-to-cream thin-bedded to massive limestone, in part with chert nodules and layers
G	TR_e^{dl}	Elika	Triassic	Light gray-to-cream dolomite
K	La	–	Triassic	Laterite
L	P_n	Nesen	Late Permian	Dark gray, well-bedded, chert-bearing limestone with interlayers of shale and dolomitic limestone

Methodology

The flowchart of the methodology used in this study is shown in Fig. 2, and consists of four phases: (1) data integration and analysis, (2) determination of the relationship between landslide occurrence and conditioning factors using the EBF model, (3) pairing EBF values to each map for running the RF model, (4) landslide susceptibility modeling using RF, and (5) validation of the landslide susceptibility map using the ROC-AUC curve (success rate and prediction rate curves).

Data integration and analysis

Generally, data collection and construction of a database of conditioning factors are the most important parts of the landslide modeling process (Ercanoglu and Gokceoglu 2002). Firstly, the landslides were identified and localized by interpreting aerial photographs, and through extensive field surveys. From the total of 153 landslides that occurred between 2005 and 2013, 70 % (105 cases) were used in the model building, while the remaining 30 % (48 cases) were used for validation according to random partition algorithm (Lee and Pradhan 2007; Bednarik et al. 2010; Ozdemir and Altural 2013; Pradhan 2013; Pourghasemi et al. 2014a; Youssef et al. 2014a, b). Failure modes of the landslides identified in the study area were determined according to the landslide classification system proposed by Varnes (1978). Most of the landslides are shallow rotational, with minimum and maximum dimensions of 350 m² and 150,420 m², respectively. Some examples of landslides

identified in the Safarood Watershed are shown in Fig. 3. For landslide susceptibility mapping in the study area, eleven conditioning factors were considered: slope angle, slope aspect, altitude, plan curvature, profile curvature, topographic wetness index, lithology, land use, distance from rivers, distance from roads, and distance from faults. One of the most important factors in landslide susceptibility mapping is topography. In this study, a digital elevation model (DEM) was created by digitizing contour lines (20 m interval) and survey base points. We used the ArcGIS TIN (Triangular Irregular Network) module, and then converted the TIN to Raster (pixel size 20 m). The contour lines and points were prepared by Iran's National Cartographic Center (NCC), and we only created the DEM from these layers. Using the DEM, slope angle, slope aspect, altitude, plan curvature, profile curvature, and topographic wetness index were extracted (Fig. 4a–e). Slope angle is one of the parameters that influence landslide occurrences. The slope map of the study area was derived from the DEM and divided into five classes: <5°, 5° to 15°, >15° to 30°, >30° to 50°, and >50° (Fig. 4a). Slope aspect is another factor that correlated with the amount of solar energy received by the area. Therefore, the slope aspect layer was selected as one of the landslide-related factors, and was categorized into nine classes: (1) flat, (2) north, (3) northeast, (4) east, (5) southeast, (6) south, (7) southwest, (8) west, and (9) northwest (Fig. 4b). The altitude map was extracted from the DEM and classified into eight equal interval classes (Lee and Pradhan 2007; Pourtaghi et al. 2014): (1) <100 m, (2) 100–500 m, (3) >500-1000 m, (4) >1000–1500 m, (5) >1500–2000 m, (6) >2000–2500 m, (7) >2500–3000 m, and (8) >3000 m
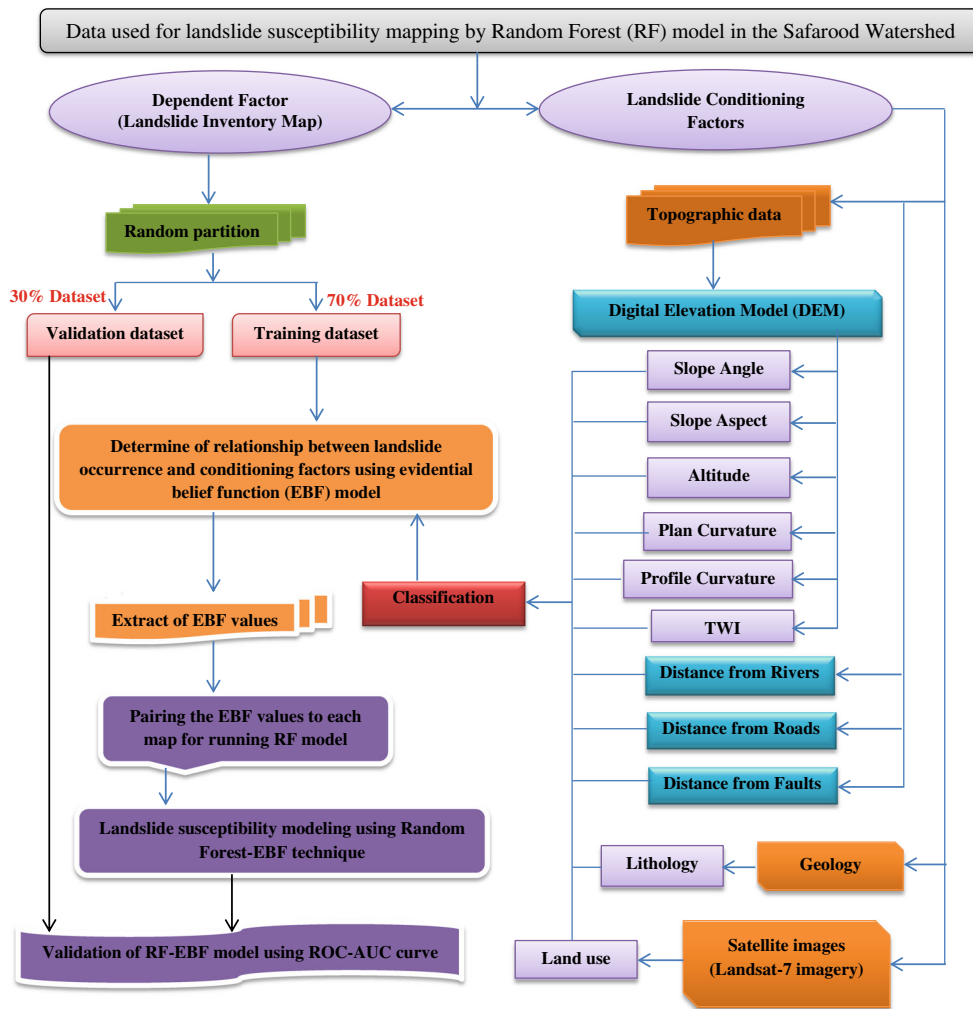
Fig. 2 Flowchart of the methodology used in this study

(Fig. 4c). Plan curvature is described as the curvature of a contour line formed by intersecting a horizontal plane with the surface. The influence of plan curvature on the slope erosion processes is the convergence or divergence of water during downhill flow (Ercanoglu and Gokceoglu 2002). The plan curvature map was produced using ArcGIS 9.3 and was classified into three categories: concave, flat, and convex (Fig. 4d). The profile curvature (also called slope profile curvature) is a primary topographic attribute. It shows the flow acceleration, deposition (positive values), and erosion (negative values) rate (Yesilnacar 2005). In addition, the profile curvature is important because it controls the velocity change of mass flowing down the slope (Talebi et al. 2007). The map was created in ArcGIS 9.3 and classified to three categories (Fig. 4e). The topographic wetness index (TWI) is defined as ln (A/tanβ), where A is upslope contributing area and β is the slope angle (Beven and Kirkby 1979). TWI has been extensively used to describe the effect of topography on the location

and size of saturated source areas of runoff generation (Beven 1997; Beven and Freer 2001). This index is commonly used to characterize the spatial distribution of soil moisture; therefore, it is used in landslide susceptibility mapping (Ozdemir 2011a, b; Oh et al. 2011a; Pourtaghi et al. 2014; Naghibi et al. 2015). The TWI map was produced using the System for Automated Geoscientific Analyses (SAGA-GIS) (Fig. 4f). The lithology map was obtained using a 1:100,000-scale geological map (GSI 1997), and the lithological units were classified into twelve groups according to lithology and its susceptibility to landslide occurrence. (Fig. 5; Table 1). The land use/land cover changes play an important role in the study of environmental issues, especially in landslide assessment (Mallick et al. 2014). The land use map was created using Landsat-7 imagery. To create the land use map, a supervised classification using the maximum likelihood algorithm was applied and verified by field survey. Nine land use classes were drawn such as dry-farming and forest

Fig. 3 Field photographs of some identified landslides in the Safarood Watershed

(DF), forest (F), forest and orchard (FO), irrigation agriculture (I), irrigation agriculture and orchard (IO), orchard (O), range (R), tea land (T), and residential area (U) (Fig. 6). Using the topographical database, the distance from rivers and distance from roads factors were calculated in ArcGIS 9.3, respectively (Fig. 7a–b). The distance from the faults was calculated at 100 m intervals, using the geological map as well (Fig. 7c).

For the classification of the conditioning factors, different methods were used, such as equal interval, natural break, and common standards such as standard division. Finally, for the application of the RF model, all conditioning factors were converted to a raster grid with 20 m × 20 m pixel size in ArcGIS 9.3 and R geostatistical packages. All the maps are in UTM (Universal Transverse Mercator) coordinate system and use the WGS84 datum (WGS84-UTM-Zone39 N).

Evidential belief function (EBF)

The Dempster–Shafer theory (Dempster 1968; Shafer 1976) is a bivariate statistical method that was applied to detect the spatial integration according to the rule of combination (Carranza 2009; Althuwaynee and Pradhan

2014). The EBF theory is based on the Dempster–Shafer rule and consists of four functions: degrees of belief (Bel), disbelief (Dis), uncertainty (Unc), and plausibility (Pls). Bel and Pls are lower and upper probabilities, respectively (Dempster 1968). Thus, Pls may be greater than or equal to Bel. Conversely, Unc and Pls are equal to Pls–Bel and 1–Unc–Bel, respectively. The degree of uncertainty indicates ignorance (or doubt) of one's belief in the proposition based on a given evidence, while the degree of disbelief is the belief that the proposition is false according to a specific evidence. If Unc = 0, then Bel = Pls. In general, Bel + Unc + Dis = 1. Dempster (1968) stated that Bel, Unc, and Dis are the functions used to integrate evidences according to combination rules. Hence, the degree of belief $\left(\text{Bel}_{C_{ij}}\right)$ is shown by Eqs. 1 and 2 (Carranza and Hale 2002):

$$\text{Bel}_{C_{ij}} = \frac{W_{C_{ij}D}}{\sum_{j=1}^{m} W_{C_{ij}D}} \tag{1}$$

$$W_{C_{ij}D} = \frac{N\left(C_{ij} \cap D\right)/N\left(C_{ij}\right)}{N(D) - N\left(C_{ij} \cap D\right)/N(T) - N\left(C_{ij}\right)} \tag{2}$$

where $N\left(C_{ij} \cap D\right)$ is density of landslide pixels that are given in D, $N\left(C_{ij}\right)$ is the total density of landslides that have occurred in the study area, $N(D)$ is the density of
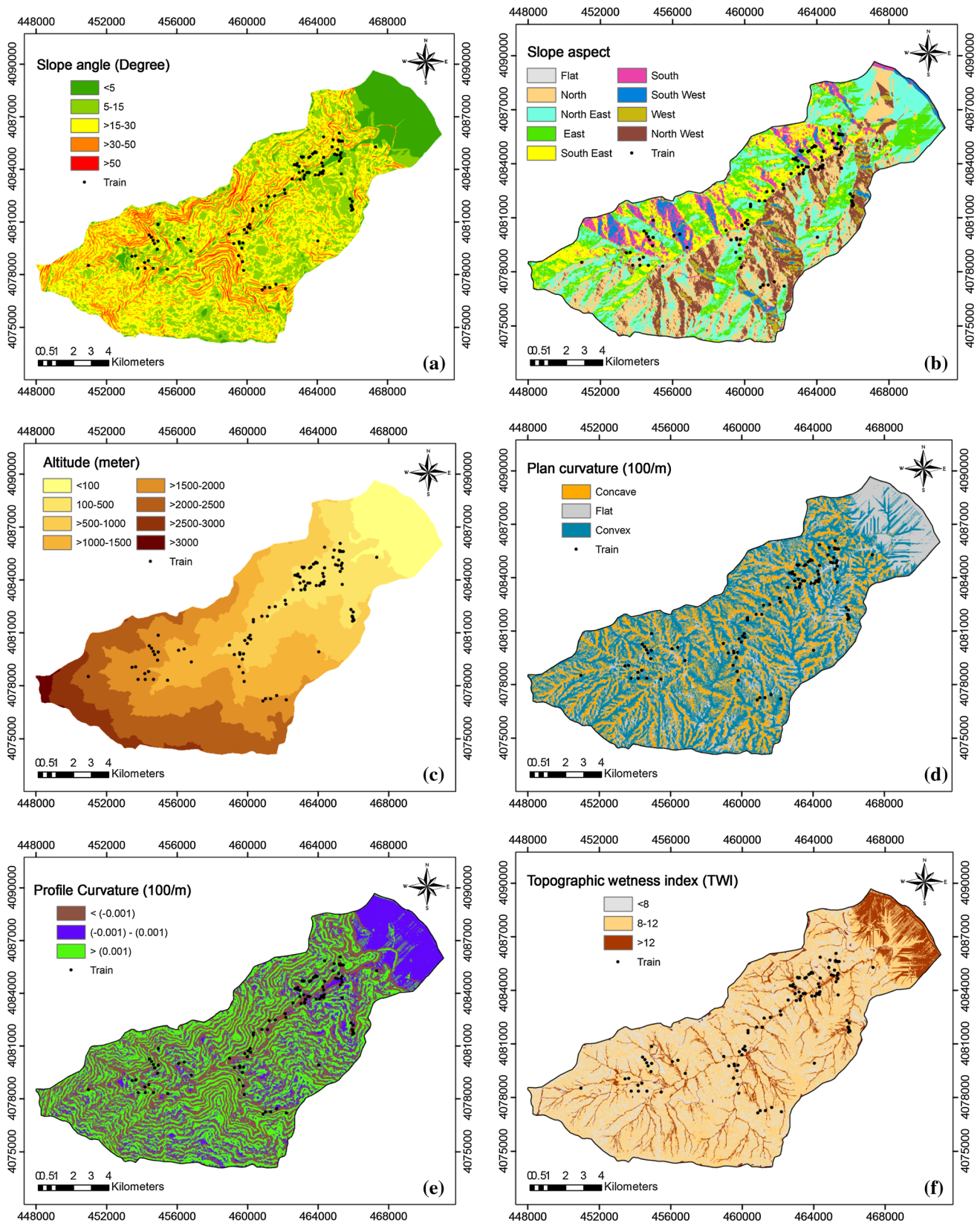
Fig. 4 Topographical variables maps of the study area: **a** slope angle, **b** slope aspect, **c** altitude, **d** plan curvature, **e** profile curvature, **f** topographic wetness index (TWI)
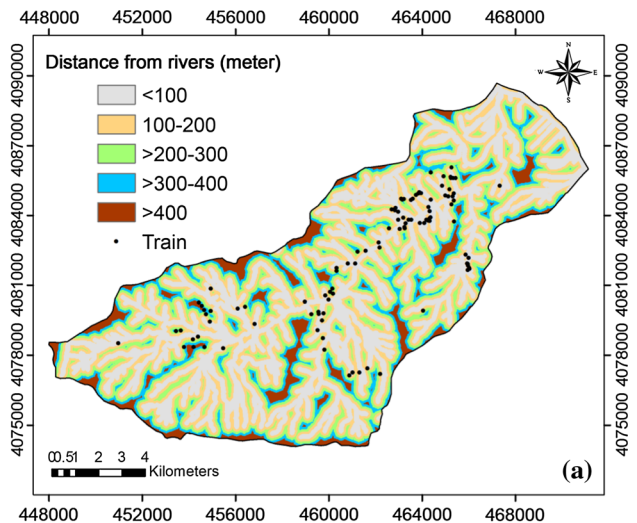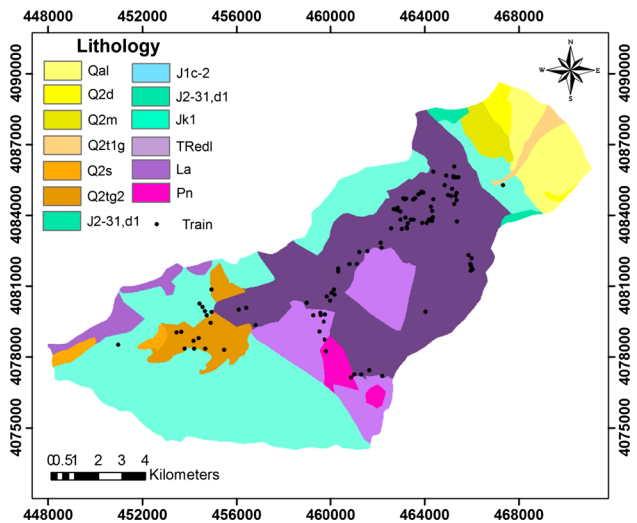
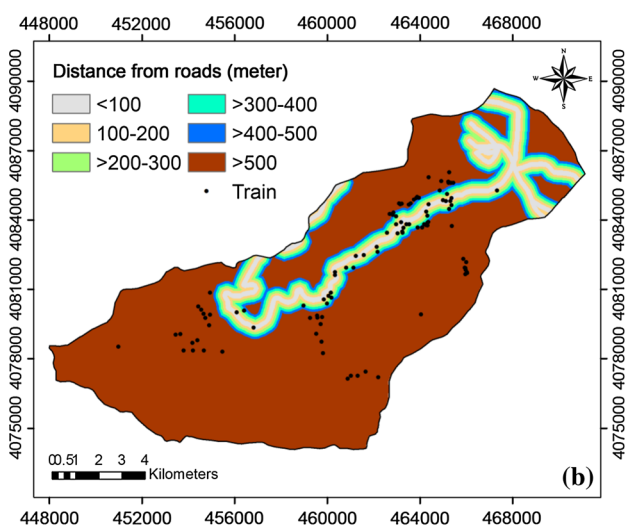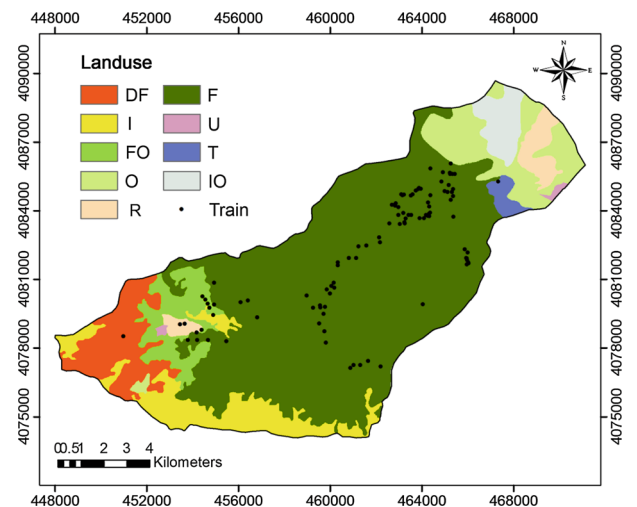**Fig. 5** The lithology map of the study area and the landslides location used for the model building

Fig. 6 The land use map of the study area. *DF* dry-farming and forest, *F* forest, *FO* forest and Orchard, *I* irrigation agriculture, *IO* irrigation agriculture and Orchard, *O* Orchard, *R* range, *T* tea land, *U* urban

pixels in *D*, and *N(T)* is the density of pixels in the whole study area *T*.

Conversely, estimation of the degree of disbelief ($Dis_{C_{ij}}$) is given by Eqs. 3 and 4 (Carranza et al. 2005; Nampak et al. 2014):

$$Dis_{C_{ij}} = \frac{W_{C_{ij}\bar{D}}}{\sum_{j=1}^{m} W_{C_{ij}\bar{D}}} \quad (3)$$

$$W_{C_{ij}\bar{D}} = \frac{N(C_{ij} \cap D)/N(C_{ij})}{N(T) - N(D) - \left[N(C_{ij}) - N(C_{ij} \cap D)\right]/N(T) - N(C_{ij})} \quad (4)$$
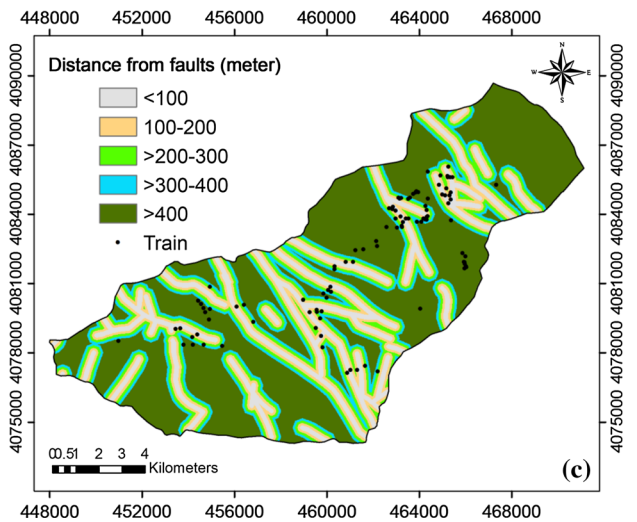

Fig. 7 Buffer maps: **a** distance from rivers, **b** distance from roads, **c** distance from faults

Accordingly, Eqs. 5 and 6 were used to calculate the uncertainty and plausibility functions.

$$\text{Unc} = \left[1 - \left(\text{Bel}_{C_{ij}} \right) - \left(\text{Dis}_{C_{ij}} \right) \right] \tag{5}$$

$$\text{Pls} = \left[1 - \left(Dis_{C_{ij}} \right) \right] \tag{6}$$

More details of the mentioned algorithm can be found in Carranza and Hale (2002), Carranza et al. (2005), and Carranza et al. (2008).

Random forests (RF)

RFs are an extension of classification and regression trees, and were first developed by Breiman (2001). RFs are widely used for data prediction and interpretation purposes. They show many appealing characteristics, such as the ability to deal with high-dimensional data, complex interactions and correlations. The RF algorithm tends to produce quite accurate models, because the ensemble reduces the instability that can be observed when building single decision trees (Williams 2011). The algorithm exploits random binary trees, which use a subset of the observations through bootstrapping techniques. For each tree grown on a bootstrap sample, the error rate for observations left out of the bootstrap sample is monitored. This is called the "out-of-bag" (OOB) error rate (Breiman 2001). Basically, the OOB accuracy indicates the accuracy of the RF predictor. It gives an estimate of test set accuracy (generalization error) (Cutler 2013). An RF tries to improve on bagging by "de-correlating" the trees. Each tree has the same expectation. In other words, the algorithm uses random feature selection at each node for the set of splitting variables (Meyer et al. 2003). This algorithm needs two original parameters to be tuned by the user: the number of trees (T) and the number of variables (m). It has been suggested (Breiman 2001; Micheletti et al. 2014) to pick a large number of trees, and the square root of the dimensionality of the input space for m (Micheletti et al. 2014). Based on two parameters, the number of trees in RF has been fixed at 1000 after an introductory analysis, and the number of variables sampled at each node was selected to be 3 to analyze the conjunct contribution of subsets of features, while maintaining a fast convergence during iterations. No calibration set is needed to tune the parameters (number of tree and number of variable) (Micheletti et al. 2014). The parameters to tune in model are iterations and learning rate. Moreover, two types of error were calculated: mean decrease in accuracy, and mean decrease in node impurity (mean decrease in Gini). The mean decrease in accuracy is determined during the calculation of the OOB error. Conversely, the mean decrease in the Gini coefficient is a measure of how much each variable contributes to the homogeneity of the nodes and leaves in the resulting RF. These importance measures can be used for ranking variables and for variable selection (Calle and Urrea 2010). Generally, mean decrease accuracy and mean decrease Gini errors in RF model have been used widely in many fields and researcher and have shown good performance for variable selection (Lawrence et al. 2006; Cutler et al. 2007; Watts et al. 2009; Stumpf and Kerle 2011; Rahmati et al. 2016), including for the classification of moderate resolution imagery that was trained with high-resolution data (Shruthi et al. 2014).

Results

Spatial relationship between landslide and conditioning factors

Results of the spatial relationship between landslide and conditioning factors using the EBF (belief, disbelief, uncertainty, and plausibility) model are shown in Table 2. For the slope degree of >50°, the belief and disbelief values were 0.341 and 0.196, respectively, which indicates a very high probability of landslide occurrence. In the case of slope aspect, the highest Bel values were related to south, southwest, and east (0.195, 0.175, and 0.140, respectively), and it showed that these categories have a positive spatial association with landslide occurrence. On the other hand, the degree of belief was lowest for other aspect categories. The relationship between landslide occurrence and altitude shows that elevations between 100 and 500 m and >500–1000 m have the highest Bel values (0.630 and 0.150, respectively), indicating that the probability of landslide occurrence in these altitudes is high. In general, the results showed that there is an inverse relationship between altitude and belief values. For plan curvature, there is a high belief and low disbelief value for convex shapes (0.349, 0.320, respectively). In the case of profile curvature, classes of <−0.001 or convex shapes have the highest belief value (0.376), followed by concave shapes (>+0.001), which mirrored findings by Gorum et al. (2011) for landslides in China. The higher (Belief degree) and lower (Disbelief degree) probabilities of landslide occurrence were obtained in areas having a TWI <8, with values of 0.359 and 0.313, respectively. In the case of lithology, there are eleven classes. The degree of belief, with respect to landslide occurrence, was higher for La (Laterite) and P_n (including dark gray, well-bedded, chert-bearing limestone with interlayers of shale and dolomitic limestone) lithological units (0.369 and 0.176, respectively), but lower or zero for other classes or units (Table 1). In the

Table 2 Spatial relationship between landslide conditioning factors and landslide locations, using the EBF model

Factors	Class	No. of pixels	No. of landslides	Belief	Disbelief	Uncertainty	Plausibility
Slope angle (°)	<5	48761	10	0.144	0.206	0.650	0.794
	5–15	114392	34	0.209	0.189	0.602	0.811
	>15–30	166085	45	0.190	0.194	0.616	0.806
	>30–50	66888	11	0.115	0.215	0.670	0.785
	>50	10286	5	0.341	0.196	0.462	0.804
Slope aspect	Flat	1900	0	0.000	0.111	0.889	0.889
	North	88416	16	0.083	0.120	0.796	0.880
	North East	85002	16	0.087	0.119	0.794	0.881
	East	69280	21	0.140	0.107	0.753	0.893
	South East	52640	20	0.175	0.103	0.722	0.897
	South	28459	12	0.195	0.106	0.700	0.894
	South West	14627	3	0.095	0.112	0.794	0.888
	West	18633	4	0.099	0.112	0.789	0.888
	North West	47455	13	0.126	0.110	0.764	0.890
Altitude (m)	<100	43223	0	0.000	0.140	0.860	0.860
	100–500	52619	58	0.630	0.064	0.306	0.936
	>500–1000	72281	19	0.150	0.124	0.725	0.876
	>1000–1500	70820	10	0.081	0.137	0.783	0.863
	>1500–2000	73974	17	0.131	0.128	0.741	0.872
	>2000–2500	70609	1	0.008	0.150	0.842	0.850
	>2500–3000	19975	0	0.000	0.131	0.869	0.869
	>3000	2911	0	0.000	0.126	0.874	0.874
Plan curvature (100/m)	Concave	164717	39	0.303	0.352	0.346	0.648
	Flat	77129	21	0.348	0.329	0.323	0.671
	Convex	164566	45	0.349	0.320	0.331	0.680
Profile curvature (100/m)	< (−0.001)	122963	35	0.376	0.320	0.303	0.60
	(−0.001)–(+0.001)	101862	20	0.260	0.362	0.378	0.638
	> (+0.001)	181587	50	0.364	0.317	0.319	0.683
TWI	<8	98166	29	0.359	0.313	0.328	0.687
	8–12	255688	64	0.295	0.365	0.340	0.635
	>12	52558	12	0.347	0.322	0.331	0.678
Lithology	$J_{2-3}^{1,d1}$	3103	0	0.000	0.084	0.916	0.916
	J_{1-2}^{c}	148670	8	0.033	0.122	0.845	0.878
	Jk^{1}	10726	0	0.000	0.086	0.914	0.914
	La	132809	79	0.369	0.031	0.600	0.969
	P_n	7048	2	0.176	0.083	0.741	0.917
	Qa^l	4111	0	0.000	0.084	0.916	0.916
	Q_2^d	24466	0	0.000	0.089	0.911	0.911
	Q_2^m	4732	0	0.000	0.085	0.915	0.915
	Q_2^{t1g}	7877	0	0.000	0.085	0.915	0.915
	Q_2^s	4047	0	0.000	0.084	0.916	0.916
	$Q_{2\ 2}^{tg}$	17624	9	0.317	0.080	0.604	0.920
	TR_e^{dl}	41199	7	0.105	0.087	0.808	0.913
Landuse	DF	29183	1	0.031	0.125	0.845	0.875
	F	259750	95	0.327	0.031	0.642	0.970
	FO	21339	4	0.167	0.119	0.714	0.881
	I	11290	3	0.237	0.117	0.646	0.883
	IO	12056	0	0.000	0.121	0.879	0.879
	O	33045	0	0.000	0.127	0.873	0.873
	R	34280	1	0.026	0.127	0.847	0.873
	T	4216	1	0.212	0.117	0.671	0.883
	U	1253	0	0.000	0.117	0.883	0.883

Table 2 continued

Factors	Class	No. of pixels	No. of landslides	Belief	Disbelief	Uncertainty	Plausibility
Distance from rivers (m)	<100	155235	49	0.310	0.174	0.516	0.826
	100–200	110944	33	0.292	0.191	0.517	0.809
	>200–300	73964	16	0.212	0.209	0.578	0.791
	>300–400	39411	6	0.149	0.211	0.640	0.789
	>400	26858	1	0.037	0.214	0.749	0.786
Distance from roads (m)	<100	20158	13	0.225	0.135	0.640	0.865
	100–200	18326	10	0.191	0.138	0.671	0.862
	>200–300	17590	7	0.139	0.143	0.718	0.857
	>300–400	15877	8	0.176	0.141	0.683	0.859
	>400–500	15325	9	0.205	0.139	0.656	0.861
	>500	319136	58	0.063	0.305	0.632	0.695
Distance from faults (m)	<100	53152	19	0.229	0.182	0.588	0.818
	100–200	52849	24	0.291	0.172	0.537	0.828
	>200–300	50466	20	0.254	0.179	0.567	0.821
	>300–400	42774	8	0.120	0.200	0.680	0.800
	>400	207171	34	0.105	0.267	0.628	0.733

DF dry-farming and forest, *F* forest, *FO* forest and orchard, *I* irrigation agriculture, *IO* irrigation agriculture and orchard, *O* orchard, *R* range, *T* tea land, *U* urban

Total pixels = 404612, Total landslides = 105

Table 3 Confusion matrix from RF model (0 = non-landslide or negative, 1 = landslide or positive)

	Predicted		Overall class error
	0	1	
			Actual
0	73	32	0.3048
1	21	84	0.200

case of land use, the degree of belief was higher for forest (0.327) and irrigation agriculture (0.237) land use types; these classes also showed lower Dis values of 0.031 and 0.117, respectively. Regarding distance from rivers, the highest belief and lowest disbelief values are found for a distance <100 m. The results indicated that landslide occurrence decreases with an increase in distance from rivers. The highest belief and lowest disbelief values in case of distance from roads were related to the class of <100 m as well. The results revealed that this class had the highest probability in landslide occurrence. In the case of distance from faults, the <200 m class had the highest belief (Bel = 0.291) and lowest disbelief (Dis = 0.172) values, respectively. The results show that with growing distance from faults landslide occurrence also increases, thus there is a direct relationship between distance from faults and landslide susceptibility. Gorum et al. (2011) also stated that landslide density decreases gradually with increasing distance from surface ruptures.

Random forests

In the RF model, 1000 trees (ntree = 1000) and just three variables (mtry = 3) were considered for the split point in each node. The out-of-bag error (OOB) was used to evaluate the performance of the model. The obtained OOB error rate was 25 %, thus the model accuracy is 75 %, which is a reasonably good model. Stumpf and Kerle (2011) reported that the model with the lowest OOB is the best. The overall measure of accuracy can be seen in the confusion matrix that records the disagreement between the

Table 4 Conditioning factors' importance based on the random forest model

Variables	0	1	MDA	MDG
Slope aspect	2.71	3.75	4.58	2.85
Altitude	7.71	23.63	21.60	11.70
Distance from faults	0.98	7.72	6.19	5.34
Land use	4.03	17.70	15.37	4.59
Lithology	8.66	24.65	23.18	9.15
Plan curvature	−0.66	−0.27	−0.65	4.02
Profile curvature	−0.83	1.20	0.24	4.25
Distance from rivers	2.77	9.45	9.04	5.74
Distance from roads	7.18	18.26	18.03	8.43
Slope angle	2.54	4.44	4.95	5.25
TWI	1.96	3.14	3.54	4.99

0 non-landslide, *1* landslide, *MDA* mean decrease in accuracy, *MDG* mean decrease in Gini

Fig. 8 Two measures of variable importance calculated by the random forest algorithm. *DEM* altitude, *Road* distance from roads, *LU* land use, *River* distance from rivers, *Fault* distance from faults, *Slope* slope angle, *Aspect* slope aspect, TWI topographic wetness index, *Prof* profile curvature, *Plan* plan curvature
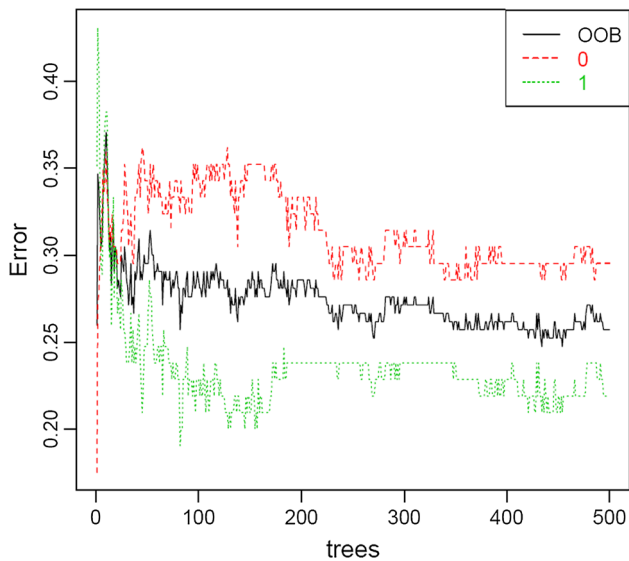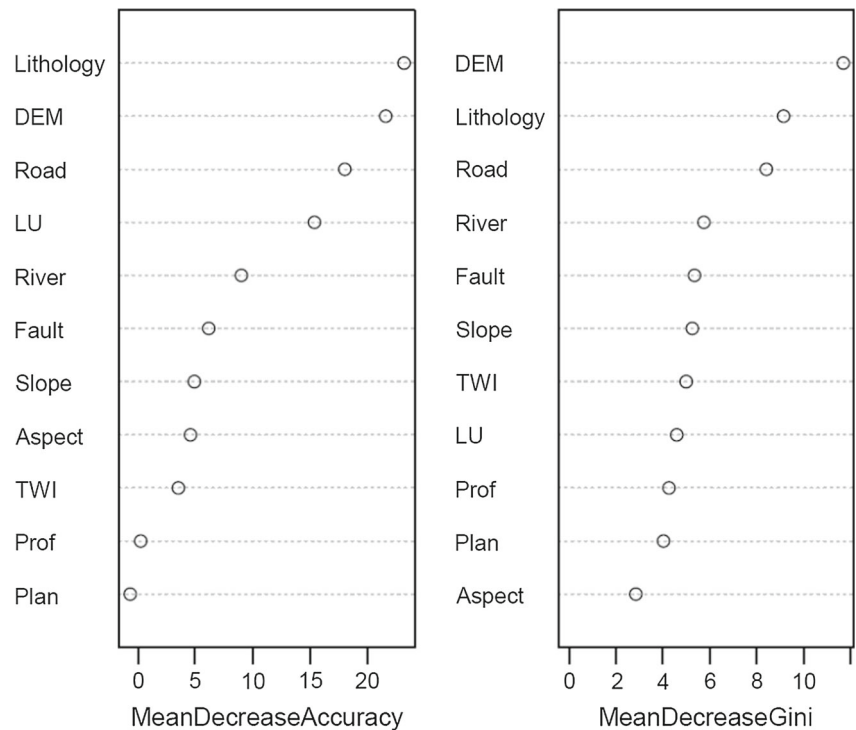

Fig. 9 OOB error plot of the random forests algorithm in the study area (*0* non-landslide, *1* landslide)

final model predictions and the actual outcomes of the training observations (Table 3). It showed that the model and training dataset agree that for 73 of the observations landslide absence is correctly predicted (69.5 %), and that of the actual 105 landslide locations 84 (80 %) were correctly predicted. One of the most important problems in RFs, compared with a single decision tree, is that it becomes quite a bit more difficult to readily understand the

discovered knowledge of 1000 trees. This was solved by assessing the relative variable importance (Table 4). The higher values in Table 4 indicate that the associated variable is relatively more important. For better understanding, a visual plot of the two measures of variable importance calculated by the RFs technique is shown in Fig. 8. Clearly, lithology is the most important variable, followed by altitude, distance from roads, and land use. The accuracy measure then includes distance from roads and land use factors as the next most important. In contrast, altitude, lithology, distance from roads, rivers, and faults had higher importance based on the Gini measure. Another key output of the RFs model is an error plot. This plot portrays the accuracy of forest of trees (*y*-axis) against the number of tress that have been included in the forest (*x*-axis; Fig. 9). It can be seen that going beyond about 50 trees in the forest adds low values (0.26–0.3), when considering the OOB error rate. The two other plots show the changes in error rate associated with the model prediction. Finally, the landslide susceptibility map (LSM) using the RF algorithm was constructed, and classified based on the natural break classification scheme in ArcGIS 9.3 (Ozdemir 2011a; Pourghasemi et al. 2012a, b; Mohammady et al. 2012; Pourghasemi et al. 2013) into low, moderate, high and very high potential classes. The landslide susceptibility map achieved from the RF method showed that about 33 % of the study area falls into the low landslide susceptibility class, while 28, 25, and 13 % were identified as having

Fig. 10 Landslide susceptibility map based on the random forests (RF) algorithm
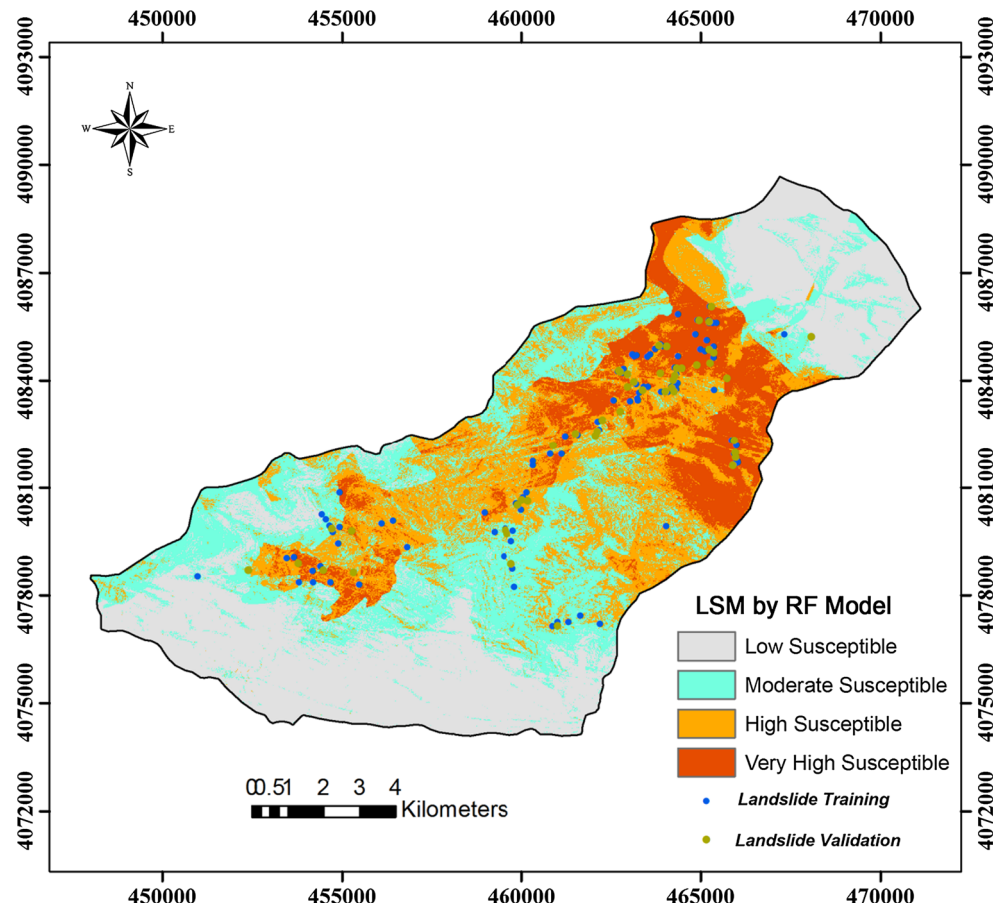
moderate, high and very high susceptibility, respectively (Fig. 10). These are substantial advantages that confine the generation of outliers, especially when working with terrain variables with a high frequency of missing data, and an intrinsic uncertainty in the assignment to the correct class also in surveyed areas. This is a typical problem for classes such as soil type, for which only a few point locations have been directly surveyed (Catani et al. 2013).

Validation of landslide susceptibility map

To determine the accuracy of the landslide susceptibility map created using the RF algorithm, the receiver operating characteristic (ROC) curve was used. This is a common method used to assess the diagnostic test accuracy. In the ROC analysis, the area under the curve (AUC) value ranges from 0.5 to 1.0, and is used to evaluate the model accuracy (Nandi and Shakoor 2010). If the model does not predict the landslide occurrences better than chance, the AUC would equal 0.5. A ROC curve of 1.0 shows perfect prediction (Yesilnacar 2005). The quantitative–qualitative relationship between AUC and prediction accuracy can be classified as follows: >0.9–1, excellent; >0.8–0.9, very good; >0.7–0.8, good; >0.6–0.7, average; and 0.5–0.6,

poor (Yesilnacar 2005). In this study, the landslide locations which were not used during the model building process were used to verify the landslide susceptibility map. For this purpose, success rate and prediction rate curves were created. The success rate method used the training landslide pixels that were used in establishing the landslide models, thus it is not a suitable method for assessing the prediction capability of the models (Bui et al. 2012). However, the method can help in determining how well the resulting landslide susceptibility maps have classified the areas of existing landslides. Another validation technique is the prediction rate curve, which explains how well the model predicts a landslide (Mohammady et al. 2012; Akgun et al. 2012; Ozdemir and Altural 2013). The AUC values of the ROC curve for RF model and the mentioned curves (success rate and prediction rate curves, respectively) were found to be 0.8520 (85.2 %) and 0.8177 (81.8 %), respectively (Fig. 11a, b). It can be concluded that this model provides reasonable results according to the success rate and prediction rate curves for landslide susceptibility mapping of the study area. Stumpf and Kerle (2011) stated that the RF algorithm provided relatively high accuracies of up to 87 % for sites in Haiti and Wenchuan. Vorpahl et al. (2012) compared eight different
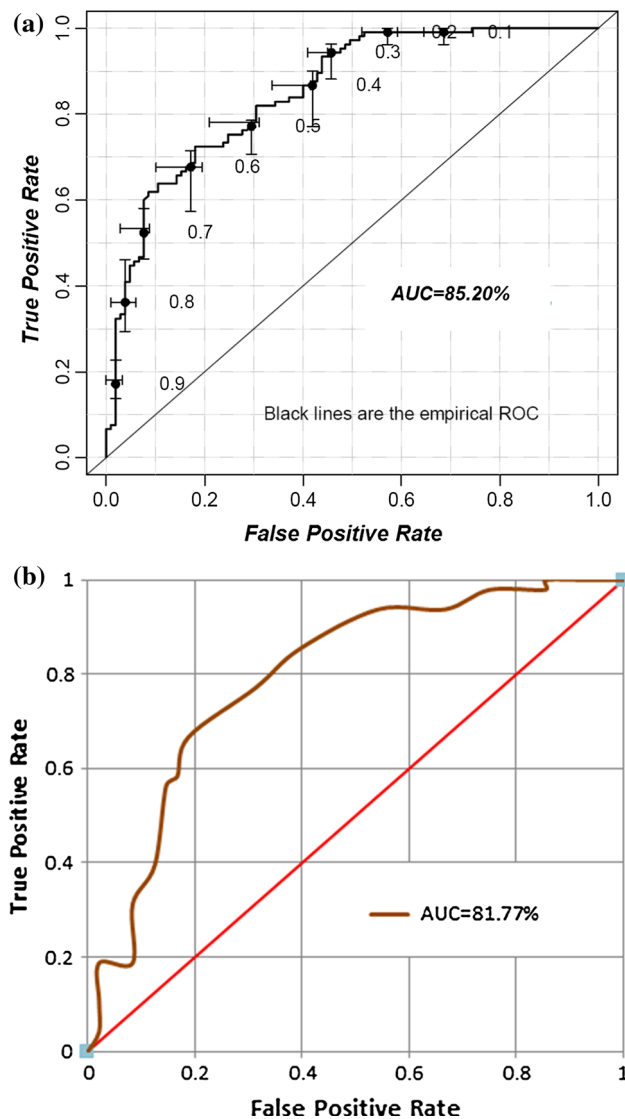
Fig. 11 ROC curves for the landslide susceptibility map produced by the random forest algorithm: **a** success rate curve **b** prediction rate

techniques for landslide susceptibility mapping in Southern Ecuador. They indicated that RF and boosted regression tree (BRT) showed the best performance in tenfold cross-validation. Meanwhile, they found that a generalized additive model (GAM) and general linear model (GLM) with stepwise backwards variable selection performed equally well. Bui et al. (2013) used an EBF-fuzzy logic hybrid model for landslide susceptibility mapping in the Vietnam. Results indicated that EBF-fuzzy logic model is better than EBF and Fuzzy models when applied separately in the study area. Esposito et al. (2014) for landslide susceptibility mapping in Rio de Janeiro, Brasil used RF and logistic regression (LR) models. Their results showed that AUC for RF and LR models were 0.81 and 0.72, respectively. Therefore, RF model shows better accuracy than LR. Trigila et al. (2015) performed a comparative study on

landslide susceptibility assessment using three models, namely logistic regression, RF, and frequency ratio. The validation of results showed that the RF model had the highest AUC with 4*4 m regular grid sampling. In another study, Goetz et al. (2015) applied different modeling techniques, such as GLM, GAM, weight of evidence, SVM, RF, and bootstrap aggregated classification tree with penalized discriminant analysis (BPLDA) for landslide susceptibility mapping in three areas in Austria. Meanwhile, the difference of AUC-ROC values ranged from 2.9 to 8.9 %, but the results showed that the RF model performed slightly better than the other models in all study areas. Youssef et al. (2015), in assessing landslides in the Wadi Tayyeh Basin, Asir Region, in Saudi Arabia, used different data mining techniques, such as RF, BRT, classification and regression tree (CART), and GLM models. The prediction rate curves indicated that RF (AUC = 81.2 %), BRT (AUC = 85.6 %), and CART (AUC = 86.2 %) have the highest AUC in comparison with generalized linear models (AUC = 76.9 %). These results are in agreement with our finding, showing that RF-EBF model has a good performance for landslide susceptibility assessment.

Conclusion

Different methods have been proposed for landslide susceptibility mapping in the literature, with the accuracy of different methods still being debated. The reliability of landslide susceptibility maps mainly depends on the amount and the quality of available data, the working scale, and the selection of the best methodology (Baeza and Corominas 2001). The RF technique has interesting features that other models lack. It can effectively handle missing data during both the training and validating steps. Because of its ensemble design, it can apply a prediction even when some of the input values are missing. Moreover, during the process of modeling, a measure of variables importance can be calculated, lending insight to the particular system being modeled (Ball 2009). Therefore, in this study a GIS-based RF model was used for landslide susceptibility mapping in the west of Mazandaran Province, northern Iran. At the same time it served as a feasibility assessment to determine if the approach can be used for other un-/under mapped parts of the country. Of the total of 153 landslide locations in the study area, 105 cases were used as training data, while the remaining 48 cases were used for validation purposes. To perform the landslide susceptibility mapping, eleven conditioning factors—slope angle, slope aspect, altitude, plan curvature, profile curvature, topographic wetness index, lithology, land use, distance from rivers, distance from roads, and distance from

faults—were considered. The EBF model was used as an initial bivariate statistical method to evaluate the correlation between the landslides and classes of each conditioning factors. Subsequently RF was applied to produce the landslide susceptibility map for the study area. For validation of the model the AUC was used. The RF model resulted in a success rate of 85.2 %, and a prediction rate of 81.8 % accuracy. In summary, based on the overall assessment, the proposed algorithm is objective and an applicable estimator that improves the predictive accuracy. It is also suitable for landslide studies and land use planning in different parts of the country, for which to date no comprehensive landslide susceptibility assessment has been done. Therefore, we propose a coupled RF and EBF analysis for landslide susceptibility in Iran, as it surpasses the performance of established methods such as bivariate and multivariate statistical models.

References

Akgun A, Sezer EA, Nefeslioglu HA, Gokceoglu C, Pradhan B (2012) An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. Comput Geosci 38:23–34

Althuwaynee O F, Pradhan B (2014) An alternative technique for landslide inventory modeling based on spatial pattern characterization. In: Rahman AA et al. (eds) Geoinformation for Informed Decisions. Lecture notes in geoinformation and cartography, pp 35–48. doi:10.1007/978-3-319-03644-1_3

Baeza C, Corominas J (2001) Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth Surf Proc Landf 26:1251–1263

Ball RL (2009) Comparison of random forest, artificial neural network, and multi-linear regression: a water temperature prediction case. In: Seventh conference on artificial intelligence and its applications to the environmental sciences, pp 1–6,

Bednarik M, Magulova B, Matys M, Marschalko M (2010) Landslide susceptibility assessment of the Kralovany-Liptovsky Mikulas railway case study. Phys Chem Earth Parts A/B/C 35(3–5):162–171

Beven KJ (1997) TOPMODEL: a critique. Hydrol Process 11:1069–1086

Beven KJ, Freer J (2001) A dynamic TOPMODEL. Hydrol Process 15:1993–2011

Beven K, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrol Sci Bull 24:43–69

Bortoloti FD, Castro Junior RM, Araújo LC, de Morais MGB (2015) Preliminary landslide susceptibility zonation using GIS-based fuzzy logic in Vitória, Brazil. Environ Earth Sci 74(3):2125–2141

Breiman L (2001) Random forests. Mach Learn 45(l):5–32

Bui TD, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in Vietnam using support vector machines, decision tree and Naive Bayes models. Math Probl Eng 2012:1–26. doi:10.1155/2012/974638

Bui DT, Pradhan B, Revhaug I, Nguyen DB, Pham HV, Bui QN (2013) A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam). Geomat Nat Hazards Risk. doi:10.1080/19475705.2013.843206

Calle ML, Urrea V (2010) Letter to the Editor: Stability of random forest importance measures. Brief Bioinform 12(1):86–89

Carranza EJM (2009) Controls on mineral deposit occurrence inferred from analysis of their spatial pattern and spatial association with geological features. Ore Geol Rev 35:383–400

Carranza EJM, Hale M (2002) Evidential belief functions for data-driven geologically constrained mapping of gold potential, Baguio district, Philippines. Ore Geol Rev 22:117–132

Carranza EJM, Woldai T, Chikambwe EM (2005) Application of data-driven evidential belief functions to prospectivity mapping for aquamarine-bearing pegmatites, Lundazi District, Zambia. Nat Resour Res 14:47–63

Carranza EJM, van Ruitenbeek FJA, Hecker C, van der Meijde M, van der Meer FD (2008) Knowledge-guided data-driven evidential belief modeling of mineral prospectivity in Cabo de Gata, SE Spain. Int J Appl Earth Obs Geoinf 10:374–387

Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat Hazards Earth Syst Sci 13:2815–2831

Chen T, Niu R, Du B, Wang Y (2014) Landslide spatial susceptibility mapping by using GIS and remote sensing techniques: a case study in Zigui County, the Three Georges reservoir, China. Environ Earth Sci 73(9):5571–5583

Constantin M, Bednarik M, Jurchescu MC, Vlaicu M (2011) Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania). Environ Earth Sci 63:397–406

Cutler A (2013) Trees and random forests. NIH 1R15AG037392-01, p 92

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in Ecology. Ecology 88(11):2783–2792

Das I, Stein A, Kerle N, Dadhwal VK (2012) Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models. Geomorphology 179:116–125

Dempster AP (1968) Generalization of Bayesian inference. J R Stat Soc Series B 30:205–247

Dou J, Yamagishi H, Pourghasemi HR, Song X, Ali YP, Xu Y, Zhu Z (2015) An integrated model for the landslide susceptibility assessment on Osado Island, Japan. Nat Hazards. doi:10.1007/s11069-015-1799-2

Ercanoglu M, Gokceoglu C (2002) Assessment of landslide susceptibility for a landslide prone area (north of Yenice, NW Turkey) by fuzzy approach. Environ Geol 41(6):720–730

Ercanoglu M, Gokceoglu C (2004) Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). Eng Geol 75:229

Esposito C, Barra A, Evans SG, Mugnozza GS, Delaney K (2014) Landslide susceptibility analysis by the comparison and integration of random forest and logistic regression methods; application to the disaster of Nova Friburgo-Rio de Janeiro, Brasil (January 2011). Geophys Res Abstr 16:11407

Fiorucci F, Antonini G, Rossi M (2015) Implementation of landslide susceptibility in the Perugia Municipal Development Plan (PRG). In: Lollino et al. (eds) Engineering geology for society and territory, vol 5, pp 769–772

Geology Survey of Iran (GSI) (1997) http://www.gsi.ir/Main/Lang_en/index.html

Goetz JN, Guthrie RH, Brenning A (2011) Integrating physical and empirical landslide susceptibility models using generalized additive models. Geomorphology 129:376–386

Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci 81:1–11

Gorum T, Fan X, van Westen CJ, Huang RQ, Xu Q, Tang C, Wang G (2011) Distribution pattern of earthquake-induced landslides triggered by the 12 May 2008 Wenchuan earthquake. Geomorphology 133:152–167

Hasekiogullari GD, Ercanoglu M (2012) A new approach to use AHP in landslide susceptibility mapping: a case study at Yenice (Karabuk, NW Turkey). Nat Hazards 63(2):1157–1179

I.R. of Iran Meteorological Organization (2012) http://www.mazandaranmet.ir/

Iranian Landslide Working Party (ILWP) (2007) Iranian landslides list. Forest, Rangeland and Watershed Association, Iran, p 60

Jaafari A, Najafi A, Pourghasemi HR, Rezaeian J, Sattarian A (2014) GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. Int J Environ Sci Technol 11:909–926

Lawrence RL, Wood SD, Sheley RL (2006) Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (Random Forest). Remote Sens Environ 100:356–362

Lee S, Choi J (2004) Landslide susceptibility mapping using GIS and the weight-of evidence model. Int J Geogr Inf Sci 18(8):789–814

Lee S, Pradhan B (2007) Landslide hazard mapping at Selangor, using frequency ratio and logistic regression models. Landslides 4:33–41

Lee MJ, Choi JW, Oh HJ, Won JS, Park I, Lee S (2012) Ensemble-based landslide susceptibility maps in Jinbu area, Korea. Environ Earth Sci 67(1):23–37

Lee J, Lee K, Joung I, Joo K, Brooks BR, Lee J (2015a) Sigma-RF: prediction of the variability of spatial restraints in template-based modeling by random forest. BMC Bioinform. doi:10.1186/s12859-015-0526-z

Lee MJ, Park I, Lee S (2015b) Forecasting and validation of landslide susceptibility using an integration of frequency ratio and neuro-fuzzy models: a case study of Seorak mountain area in Korea. Environ Earth Sci 74(1):413–429

Li XZ, Kong JM (2014) Application of GA–SVM method with parameter optimization for landslide development prediction. Nat Hazards Earth Syst Sci 14:525–533

Li C, Ma T, Sun L, Li W, Zheng A (2012) Application and verification of a fractal approach to landslide susceptibility mapping. Nat Hazards 61(1):169–185

Mallick J, Al-Wadi H, Atiqur Rahman, Ahmed M (2014) Landscape dynamic characteristics using satellite data from a mountainous watershed of Abha, Kingdom of Saudi Arabia. Environ Earth Sci. doi:10.1007/s12665-014-3408-1

Martha TR, Kerle N, van Westen CJ, Jetten VG, Kumar KV (2012) Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories. ISPRS J Photogram Remote Sens 67:105–119

Meten M, Bhandary NP, Yatabe R (2015) Application of GIS-based fuzzy logic and rock engineering system (RES) approaches for landslide susceptibility mapping in Selelkula area of the Lower Jema River Gorge, Central Ethiopia. Environ Earth Sci 74(4):3395–3416

Meyer D, Leisch F, Hornik K (2003) The support vector machine under test. Neurocomputing 55(1–2):169–186

Micheletti N, Foresti L, Robert S, Leuenberger M, Pedrazzini A, Jaboyedoff M, Kanevski M (2014) Machine learning feature selection methods for landslide susceptibility mapping. Math Geosci 46:33–57

Mihir M, Malamud B, Rossi M, Reichenbach P, Ardizzone F (2014) Landslide susceptibility statistical methods: a critical and systematic literature review. Geophys Res Abstr 16(EGU2014-9814):2014

Miner AS, Vamplew P, Windle DJ, Flentje P, Warner P (2010) A comparative study of various data mining techniques as applied to the modeling of landslide susceptibility on the Bellarine Peninsula, Victoria, Australia. Geologically active. In: Proceedings of the 11th IAEG congress of the international association of engineering geology and the environment, Auckland, New Zealand

Mohammady M, Pourghasemi HR, Pradhan B (2012) Landslide susceptibility mapping at Golestan Province, Iran: a comparison between frequency ratio, Dempster-Shafer, and weights-of-evidence models. J Asian Earth Sci 61:221–236

Moosavi V, Niazi Y (2015) Development of hybrid wavelet packet-statistical models (WP-SM) for landslide susceptibility mapping. Landslides. doi:10.1007/s10346-014-0547-0

Naghibi SA, Pourghasemi HR, Pourtaghi ZS, Rezaei A (2015) Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan Watershed, Iran. Earth Sci Inform 8(1):171–186

Nampak H, Pradhan B, Manap MA (2014) Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. J Hydrol 513:283–300

Nandi A, Shakoor A (2010) A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. Eng Geol 110:11–20

Nourani V, Pradhan B, Ghaffari H, Sharifi SS (2014) Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. Nat Hazards 71(1):523–547

Oh HJ, Park NW, Lee SS, Lee S (2011a) Extraction of landslide-related factors from ASTER imagery and its application to landslide susceptibility mapping. Int J Remote Sens 33(10):3211–3231

Oh HJ, Kim YS, Choi JK et al (2011b) GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. J Hydrol 399:158–172

Osna T, Sezer EA, Akgun A (2014) GeoFIS: an integrated tool for the assessment of landslide susceptibility. Comput Geosci 66:20–30

Ozdemir A (2011a) Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). J Hydrol 405:123–136

Ozdemir A (2011b) GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. J Hydrol 411:290–308

Ozdemir A, Altural T (2013) A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. J Asian Earth Sci 64:180–197

Park NW (2014) Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets. Environ Earth Sci 73(3):937–949

Park I, Lee S (2014) Spatial prediction of landslide susceptibility using a decision tree approach: a case study of the Pyeongchang area, Korea. Int J Remote Sens. doi:10.1080/01431161.2014.943326

Park S, Choi C, Kim B, Kim J (2013) Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environ Earth Sci 68(5):1443–1464

Poiraud A (2014) Landslide susceptibility-certainty mapping by a multi-method approach: a case study in the Tertiary basin of Puy-en-Velay (Massif central, France). Geomorphology. doi:10.1016/j.geomorph.2014.04.001

Pourghasemi HR, Beheshtirad M (2014) Assessment of a data-driven evidential belief function model and GIS for groundwater

potential mapping in the Koohrang Watershed, Iran. Geocarto Int. doi:10.1080/10106049.2014.966161

Pourghasemi HR, Mohammady M, Pradhan B (2012a) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena 97:71–84

Pourghasemi HR, Pradhan B, Gokceoglu C (2012b) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Nat Hazards 63(2):965–996

Pourghasemi HR, Moradi HR, Fatemi Aghda SM (2013) Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. Nat Hazards 69:749–779

Pourghasemi HR, Moradi HR, Fatemi Aghda SM (2014a) GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). Arab J Geosci 7(5):1857–1878

Pourghasemi HR, Moradi HR, Fatemi Aghda SM, Sezer EA, Goli Jirandeh A, Pradhan B (2014b) Assessment of fractal dimension and geometrical characteristics of landslides identified in North of Tehran, Iran. Environ Earth Sci 71:3617–3626

Pourtaghi Z, Pourghasemi HR, Rossi M (2014) Forest fire suscep- tibility mapping in the Minudasht Forests, Golestan Province, Iran. Environ Earth Scis. doi:10.1007/s12665-014-3502-4

Pradhan B (2010) Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. J Indian Soc Remote Sens 38(2):301–320

Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci 51:350–365

Pradhan B, Buchroithner MF (2010) Comparison and validation of landslide susceptibility maps using an artificial neural network model for three test areas in Malaysia. Environ Eng Geosci 16(2):107–126

Rahmati O, Pourghasemi HR, Melesse A (2016) Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. Catena 137:360–372

Ren F, Wu X, Zhang K, Niu R (2014) Application of wavelet analysis and a particle swarm-optimized support vector machine to predict the displacement of the Shuping landslide in the Three Gorges, China. Environ Earth Sci 73(8):4791–4804

Rodriguez-Galiano V, Chica-Rivas M (2012) Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and digital terrain models. Int J Digit Earth 7(6):1–18

Sezer EA, Pradhan B, Gokceoglu C (2011) Manifestation of an adaptive neuro-fuzzy model on landslide susceptibility mapping: Klang valley, Malaysia. Expert Syst Appl 38(7):8208–8219

Shafer G (1976) A mathematical theory of evidence, vol 1. Princeton University, Princeton

Shruthi RBV, Kerle N, Jetten VG, Stein A (2014) Object-based gully system prediction from medium resolution imagery using random forests. Geomorphology 216:283–294

Song Y, Gong J, Gao S, Wang D, Cui T, Li Y, Wei B (2012) Susceptibility assessment of earthquake induced landslides using Bayesian network: a case study in Beichuan, China. Comput Geosci 42:189–199

Stumpf A, Kerle N (2011) Object-oriented mapping of landslides using random forests. Remote Sens Environ 115:2564–2577

Su C, Wang L, Wang X, Huang Z, Zhang X (2015) Mapping of rainfall-induced landslide susceptibility in Wencheng, China,

using support vector machine. Nat Hazards. doi:10.1007/s11069-014-1562-0

Talebi A, Uijlenhoet R, Troch PA (2007) Soil moisture storage and hillslope stability. Nat Hazards Earth Syst Sci 7:523–534

Trigila A, Carla I, Carlo E, Gabriele SM (2015) Comparison of logistic regression and random forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). Geomorphology. doi:10.1016/j.geomorph.2015.06.001

Umar Z, Pradhan B, Ahmad A, Jebur MN, Tehrany MS (2014) Earthquake-induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. Catena 118:124–135

Varnes DJ (1978) Slope movement types and processes. In: Schuster RL, Krizek RJ (Eds) Landslides analysis and control. Special Report, vol. 176. Transportation Research Board, National Academy of Sciences, New York, pp 12–33

Vorpahl P, Elsenbeer H, Märker M, Schröder B (2012) How can statistical models help to determine driving factors of landslides? Ecol Model 239:27–39

Wang HB, Sassa K (2005) Comparative evaluation of landslide susceptibility in Minamata area, Japan. Environ Geol 47:956–966

Wang YT, Seijmonsbergen AC, Bouten W, Chen QT (2015) Using statistical learning algorithms in regional landslide susceptibility zonation with limited landslide field data. J Mount Sci. doi:10.1007/s11629-014-3134-x

Watts JD, Lawrence RL, Miller PR, Montagne C (2009) Monitoring of cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing. Remote Sens Environ 113:1843–1852

Weirich F, Blesius L (2007) Comparison of satellite and air photo based landslide susceptibility maps. Geomorphology 87(4):352–364

Williams G (2011) Data mining with Rattle and R. pp 245–268

Wu X, Niu R, Ren R, Peng L (2013) Landslide susceptibility mapping using rough sets and back-propagation neural networks in the Three Gorges, China. Environ Earth Sci 70(3):1307–1318

Wu XL, Ren F, Niu RQ (2014) Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. Environ Earth Sci 71(11):4725–4738

Yesilnacar EK (2005) The application of computational intelligence to landslide susceptibility mapping in Turkey, PhD Thesis. Department of Geomatics the University of Melbourne, p 423

Youssef AM (2015) Landslide susceptibility delineation in the Ar-Rayth area, Jizan, Kingdom of Saudi Arabia, using analytical hierarchy process, frequency ratio, and logistic regression models. Environ Earth Sci 73(12):8499–8518

Youssef AM, Pradhan B, Jebur MN, El-Harbi HM (2014a) Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area, Saudi Arabia. Environ Earth Sci. doi:10.1007/s12665-014-3661-3

Youssef AM, Al-kathery M, Pradhan B (2014b) Landslide suscep- tibility mapping at Al-Hasher Area, Jizan (Saudi Arabia) using GIS-based frequency ratio and index of entropy models. Geosci J. doi:10.1007/s12303-014-0032-8

Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2015) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides. doi:10.1007/s10346-015-0614-1